

[First Hit](#) [Previous Doc](#) [Next Doc](#) [Go to Doc#](#)**End of Result Set** [Generate Collection](#) [Print](#)

L5: Entry 1 of 1

File: JPAB

May 9, 2003

PUB-NO: JP02003131910A

DOCUMENT-IDENTIFIER: JP 2003131910 A

TITLE: APPARATUS, METHOD AND PROGRAM FOR DATABASE CONSTRUCTION, AND RECORDING MEDIUM

PUBN-DATE: May 9, 2003

INVENTOR-INFORMATION:

NAME	COUNTRY
TANAKA, JUN	
OKUDA, TATSUGO	
NAGAYAMA, SATORU	

ASSIGNEE-INFORMATION:

NAME	COUNTRY
TOPPAN PRINTING CO LTD	

APPL-NO: JP2001328330

APPL-DATE: October 25, 2001

INT-CL (IPC): G06F 12/00

ABSTRACT:

PROBLEM TO BE SOLVED: To provide a system for database construction, capable of constructing a database by creating data configuration files from a layout configuration files.

SOLUTION: The system for database construction constructing the database from an electronic document, in which the elements composing of the electronic document are pre-laid out comprises a defining means for rule of element name setting rule information for setting elements extracted from the electronic document, a specifying means for the rule of element name setting item information which is to serve as items of the database to the elements extracted from the electronic document, and a configuration element extracting means extracting the elements from the electronic document based on the rule information specified by the defining means and outputting the extracted elements as database information, by making the extracted elements associated with the item information set by the specifying means.

COPYRIGHT: (C)2003, JPO

[Previous Doc](#) [Next Doc](#) [Go to Doc#](#)

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2003-131910

(P2003-131910A)

(43)公開日 平成15年5月9日 (2003.5.9)

(51)Int.Cl.⁷

G 0 6 F 12/00

識別記号

5 1 3

F I

G 0 6 F 12/00

マーク (参考)

5 1 3 A 5 B 0 8 2

審査請求 未請求 請求項の数6 OL (全13頁)

(21)出願番号 特願2001-328330(P2001-328330)

(22)出願日 平成13年10月25日 (2001.10.25)

(71)出願人 000003193

凸版印刷株式会社

東京都台東区台東1丁目5番1号

(72)発明者 田中 潤

東京都台東区台東1丁目5番1号 凸版印刷株式会社内

(72)発明者 奥田 竜吾

東京都台東区台東1丁目5番1号 凸版印刷株式会社内

(74)代理人 100064908

弁理士 志賀 正武 (外6名)

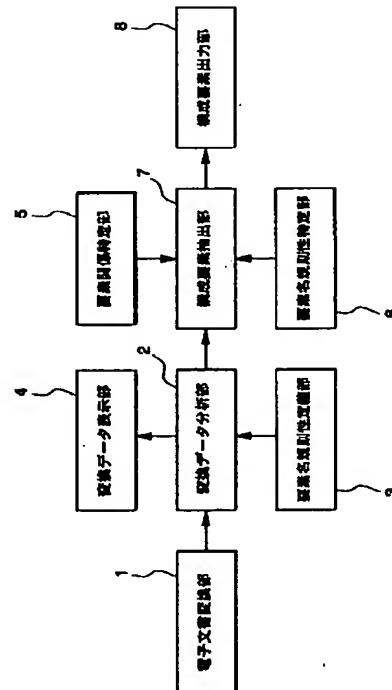
最終頁に続く

(54)【発明の名称】 データベース構築装置、データベース構築方法、データベース構築プログラム、記録媒体

(57)【要約】

【課題】 レイアウト構造ファイルからデータ構造ファイルを生成してデータベースを構築することができるデータベース構築システムを提供する。

【解決手段】 電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築システムであって、電子文書から抽出する要素を指定するためのルール情報を設定する要素名規則性定義手段と、電子文書から抽出される要素にデータベースの項目となる項目情報を設定する要素名規則性特定手段と、要素名規則性定義手段によって指定されるルール情報に基づいて、電子文書から要素を抽出し、抽出した要素を要素名規則性特定手段によって設定される項目情報を対応付けてデータベース情報として出力する構成要素抽出手段とを有する。



【特許請求の範囲】

【請求項1】 電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築システムであって、
前記電子文書から抽出する要素を指定するためのルール情報を設定またはルール情報を予め記憶するルールデータベースから読み出す要素名規則性定義手段と、
前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定する要素名規則性特定手段と、
前記要素名規則性定義手段によって指定されるルール情報に基づいて、前記電子文書から要素を抽出し、抽出した要素を前記要素名規則性特定手段によって設定される項目情報を対応付けてデータベース情報として出力する構成要素抽出手段と、
を有することを特徴とするデータベース構築装置。

【請求項2】 前記要素のうち、複数の要素を1つのグループとして関連付けするリンク情報を生成する要素関係特定手段を有し、

前記構成要素抽出手段は、前記要素関係特定手段によって生成されたリンク情報に基づいて前記電子文書から抽出される要素を関連付けるとともに、前記項目情報を対応付けてデータベース情報として出力することを特徴とする請求項1記載のデータベース構築装置。

【請求項3】 前記電子文書は、商品のカタログに関する情報を含むことを特徴とする請求項1または請求項2記載のデータベース構築装置。

【請求項4】 電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築方法であって、

前記電子文書から抽出する要素を指定するためのルール情報を設定し、

前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定し、

前記ルール情報に基づいて、前記電子文書から要素を抽出し、

抽出した要素を前記項目情報に基づき、前記要素と前記項目とを対応付けてデータベース情報として出力する、ことを特徴とするデータベース構築方法。

【請求項5】 電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築プログラムであって、

前記電子文書から抽出する要素を指定するためのルール情報を設定するステップと、

前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定するステップと、

前記ルール情報に基づいて、前記電子文書から要素を抽出するステップと、

抽出した要素を前記項目情報に基づき、前記要素と前記項目とを対応付けてデータベース情報として出力するステップと、

をコンピュータに実行させることを特徴とするデータベース構築プログラム。

【請求項6】 電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築プログラムを記録したコンピュータ読み取り可能な記録媒体であって、

前記電子文書から抽出する要素を指定するためのルール情報を設定するステップと、

前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定するステップと、

前記ルール情報に基づいて、前記電子文書から要素を抽出するステップと、抽出した要素を前記項目情報に基づき、前記要素と前記項目とを対応付けてデータベース情報として出力するステップと、

をコンピュータに実行させるデータベース構築プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、データベースの構築および更新を簡単に行うことができるデータベース構築装置、データベース構築方法、データベース構築プログラム、記録媒体に関するものである。

【0002】

【従来の技術】従来から、カタログ作成依頼者（以下、依頼者と称する）からカタログを作成依頼された場合に、印刷業者は、DTP（Desktop Publishing）が適用されたシステムを利用して、カタログを作成することが行われている。例えば、カタログに掲載するための商品の画像や、商品の名称、型式、寸法、製品紹介文などの文書等の商品情報が商品データベースに予め記憶されている。この商品データベースからDTP用コンピュータを利用して、商品情報を読み出し、画像や文書のレイアウトを決め、編集を行い、見本を印刷する。そして、印刷された見本を依頼者に確認（レイアウト、誤植の有無等）してもらい、訂正内容があれば、見本に訂正内容を記入してもらう。印刷業者は、見本に記入された訂正内容に応じて、DTP用コンピュータによってDTP上のデータを修正する。そして、修正後の見本の印刷物を依頼者に再度確認してもらい、変更がなければ、印刷し、カタログを作成する。

【0003】一方、上述した商品データベースは、次回のカタログ制作時においては、新製品などの情報が追加され、引き続き利用される。従って、依頼者によって見本に記入された訂正内容に応じて、印刷業者は、DTP用コンピュータによってDTP上のデータが修正された場合、この修正内容を商品データベースに反映させ、商品データベースを更新する必要があるが、この商品データベースの更新作業は、印刷業者が見本に記入された訂正内容に基づいて、商品データベース用のコンピュータから商品データベースの更新作業を行っていた。

【0004】

【発明が解決しようとする課題】しかしながら、従来技術によれば、印刷業者は、依頼者によって記入された見本の訂正内容に基づいて、DTP上のデータの更新作業を行うとともに、商品データベースの更新作業を行うすなわち、二重に入力作業を行う必要があり、これらの更新作業は、印刷業者にとって大きな負担であるとともに、更新作業に多大な時間がかかってしまっていた。また、商品データベースに対しても、DTP上のデータの更新と同じ更新内容を反映させる必要があるが、作業者のミスなどにより、商品データベースに確実な更新作業を行うことができない場合もあった。

【0005】本発明は、このような事情に鑑みてなされたもので、その目的は、レイアウト構造ファイルからデータ構造ファイルを生成してデータベースを構築することができるデータベース構築装置、データベース構築方法、データベース構築プログラム、記録媒体を提供することにある。

【0006】

【課題を解決するための手段】上記目的を達成するため、本発明は、電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築システムであって、前記電子文書から抽出する要素を指定するためのルール情報を設定またはルール情報を予め記憶するルールデータベースから読み出す要素名規則性定義手段と、前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定する要素名規則性特定手段と、前記要素名規則性定義手段によって指定されるルール情報に基づいて、前記電子文書から要素を抽出し、抽出した要素を前記要素名規則性特定手段によって設定される項目情報を対応付けてデータベース情報として出力する構成要素抽出手段と、を有することを特徴とする。

【0007】また、本発明は、上述のデータベース構築装置において、前記要素のうち、複数の要素を1つのグループとして関連付けするリンク情報を生成する要素関係特定手段を有し、前記構成要素抽出手段は、前記要素関係特定手段によって生成されたリンク情報に基づいて前記電子文書から抽出される要素を関連付けるとともに、前記項目情報を対応付けてデータベース情報として出力することを特徴とする。また、本発明は、上述のデータベース構築装置において、前記電子文書は、商品のカタログに関する情報を含むことを特徴とする。

【0008】また、本発明は、電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築方法であって、前記電子文書から抽出する要素を指定するためのルール情報を設定し、前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定し、前記ルール情報に基づいて、前記電子文書から要素を抽出し、抽出した要素を前記項目

10

20

30

40

50

情報に基づき、前記要素と前記項目とを対応付けてデータベース情報として出力することを特徴とする。

【0009】また、本発明は、電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築プログラムであって、前記電子文書から抽出する要素を指定するためのルール情報を設定するステップと、前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定するステップと、前記ルール情報に基づいて、前記電子文書から要素を抽出するステップと、抽出した要素を前記項目情報に基づき、前記要素と前記項目とを対応付けてデータベース情報として出力するステップと、をコンピュータに実行させることを特徴とする。

【0010】また、本発明は、電子文書を構成する要素が予めレイアウトされた電子文書からデータベースを構築するデータベース構築プログラムを記録したコンピュータ読み取り可能な記録媒体であって、前記電子文書から抽出する要素を指定するためのルール情報を設定するステップと、前記電子文書から抽出される要素にデータベースの項目となる項目情報を設定するステップと、前記ルール情報に基づいて、前記電子文書から要素を抽出するステップと、抽出した要素を前記項目情報に基づき、前記要素と前記項目とを対応付けてデータベース情報として出力するステップと、をコンピュータに実行させるデータベース構築プログラムを記録したことを特徴とする。

【0011】

【発明の実施の形態】以下、本発明の一実施形態によるデータベース構築装置を図面を参照して説明する。この実施形態においては、図2に示すような商品カタログのレイアウト構造ファイルである電子文書からデータ構造ファイルを生成し、データベースを構築する場合について説明する。ここでいうデータ構造ファイルとは、要素に意味を持たせ、その意味を指定し、表現することができるデータを記憶するファイルであり、例えば、DB、XML (eXtensible Markup Language) 等がある。レイアウト構造ファイルとは、要素を表現する位置 (レイアウト) を指定し表現することができるデータを記憶するためのファイルであり、例えば、DTPデータ、HTML (HyperText Markup Language) 等がある。図1は、この発明の一実施形態によるデータベース構築装置の構成を示す概略ブロック図である。この図において、電子文書変換部1は、外部の装置から出力される電子ファイルをデータベース構築装置において処理可能なファイル形式のデータに変換し、変換されたデータを変換データ分析部2に出力する。例えば、電子文書変換部1は、外部から送信されるDTPドキュメントデータを中間ファイルに変換する。

【0012】ここで、中間ファイルとは、レイアウト構

造ファイルであり、文字や画像、図形などが紙や画面に出力される際の体裁情報をもつ、印刷用データ、ワープロデータ、マークアップ言語データ（HTML（Hypertext Markup Language）、XML（Extensible Markup Language）、SGML（Standard Generalized Markup Language））等の形式のデータである。

【0013】変換データ分析部2は、電子文書変換部1から出力される中間ファイルに基づく画像を変換データ表示部4の画面上に出力する。この変換データ表示部4の画面上には、例えば、図2に示すような書類をまとめた事務用ファイルのカタログに関する情報が表示される。また、変換データ分析部2は、電子文書変換部1から出力される中間ファイルと要素名規則性定義部3から出力されるルール情報を構成要素抽出部7に出力する。

【0014】要素名規則性定義部3は、電子文書から抽出する要素を指定するためのルール情報を設定する。このルール情報の一例を図3に示す。この図に示すように、ルール情報のうち、文字に対するルール情報は、例えば、文字の書体名、サイズ、色、変形、字送りのうち少なくとも1つの条件を含む文字体裁に関する条件と、行頭と行末とのうち一方又は両方を含む文字が配置された文字位置に関する条件と、前後の文字または他の要素との関係を指定する関係前後関係に関する条件と、内部に含まれる文字列を指定する条件である包含文字に関する条件と、のうち、少なくともいずれか1つの条件を含む。また、抽出する対象が文字である場合、画像に対するルール情報は、ファイル名とファイル種別との一方または両方の条件を含むファイル情報を関する条件と、画像の大きさと解像度との一方または両方の条件を含む画像属性に関する条件と、画像の回転と倍率との一方または両方の条件を含むレイアウト属性に関する条件と、のうち少なくとも1つの条件を含む。また、文字と画像とに対する共通の条件となるルール情報は、レイアウト領域の大きさと枠線との一方または両方を含むレイアウト領域に関する条件を含む。また、要素名規則性定義部3は、ルール情報を予め記憶するルールデータベースを有し、必要に応じて、ルールデータベースからルール情報を読み出す。

【0015】また、ここでいう要素とは、ヘッダーやワッター、見出しや本文、挿絵など文書を構成している内容となる情報であり、商品カタログの場合は、商品の品番や価格などの商品スペックの他、写真や図形、ロゴマークなどが該当する。なお、このルール情報は、データベース構築装置の外部に接続されるキーボードやマウス等の入力デバイスから、作業者によって入力される。

【0016】変換データ表示部4は、表示装置であり、例えば、CRT（Cathode Ray Tube）や

液晶表示装置等が用いられる。

【0017】要素関係特定部5は、入力デバイスを介して入力される作業者からの指示に基づいて、各要素間を1つのグループとして関連付けするリンク情報を生成する。このリンク情報とは、例えば、図4の符号（a）、符号（b）、符号（c）、符号（d）に示す各レイアウト領域内の文字や画像の情報が1つの商品について説明するためのデータであることの関連づけを行う情報である。これにより、画像や文字がどの商品であるかが商品毎に関連づけされる。なお、ここでいうレイアウト領域とは、1つの画像または1つの文字列などが設定される領域である。

【0018】要素名規則性特定部6は、電子文書から抽出される文字や画像などの要素にデータベースの項目となる項目情報を設定する。この項目には、例えば、「品番」、「型式」、「寸法」、「価格」等がある。

【0019】構成要素抽出部7は、要素名規則性定義部3によって指定されるルール情報に基づいて、電子文書から要素を抽出し、抽出した要素を要素名規則性特定部6によって設定される項目情報を対応付けてデータベース情報として構成要素出力部8に出力する。また、構成要素抽出部7は、要素関係特定部5によって生成されたリンク情報に基づいて、電子文書から抽出される要素を関連付けるとともに、項目情報を対応付けてデータベース情報として出力する。

【0020】構成要素出力部8は、表示装置であり、構成要素抽出部7から出力されるデータベース情報を画面上に出力する。構成要素出力部8は、変換データ表示部4と共通の表示装置を利用するようにもよい。また、構成要素出力部8は、データ出力装置であってもよい。

【0021】次に、図1に示すデータベース構築装置の動作について図5から図8のフローチャートを用いて説明する。まず、図5において、外部からDTPデータの電子文書が入力されると（ステップA1）、電子文書変換部1は、入力された電子文書を中間ファイルに変換し（ステップA2）、変換後の中間ファイルを変換データ分析部2に出力する（ステップA3）。

【0022】次に、図6において、変換されたデータである中間ファイルが入力されると（ステップB1）、データベース構築装置は、構成要素の特定を行い（ステップB2）、特定された構成要素の抽出を行い（ステップB3）、抽出された構成要素をデータベース情報として構成要素出力部8から出力する（ステップB4）。

【0023】次に、図7を用いて、図6におけるステップB2の構成要素の特定処理について説明する。変換データ分析部2は、電子文書変換部1から出力された中間ファイルを変換データ表示部4に出力し、画面上に表示させる（ステップB21）。このとき、変換データ表示部4の画面上には、例えば、図2のような表示がなされ

る。

【0024】次に、要素名規則性定義部3は、作業者から入力デバイスを介して、ルール情報が入力されると、入力されたルール情報を設定する。次に、要素関係特定部5は、作業者から入力デバイスを介して各レイアウト領域間の構成要素が選択され、要素関係の関連付けが指示されると、選択された構成要素間の関連付けを行う(ステップB23)。そして、要素名規則性特定部6は、作業者から入力デバイスを介して各レイアウト領域に対して要素名が指定されると、指定された要素名が、各レイアウト領域にデータベース情報の項目として設定するとともに、作業者から入力デバイスを介して入力される指示に基づいて、要素名規則性定義部3から入力されたルール情報と項目の対応付けを行う(ステップB24)。この設定は、すべてのレイアウト領域に対して行われると、終了する(ステップB25)。

【0025】次に、図8を用いて図7のステップB24について説明する。要素関係特定部5は、作業者から入力デバイスを介して、各レイアウト領域に対して要素名が定義されると(ステップB241)、要素名規則性定義情報3から入力されたルール情報と項目の対応付けを行う。ここでは、構成要素名の定義対象が画像である場合(ステップB242)、画像情報を特定するルール情報が定義され(ステップB243)、定義対象が文字である場合、(ステップB242)、文字情報を特定するルール情報が定義される(ステップB244)。そして、抽出する対象のデータのルール情報が商品の画像、型式、寸法などの項目に対してすべて設定されると、構成要素名の定義が終了する(ステップB245)。

【0026】そして、構成要素抽出部7は、変換データ分析部2から出力される中間ファイル内のデータから、要素関係特定部5からの指示に基づいて各レイアウト領域間の関連づけを行い、関連づけされた各レイアウト領域内のデータに対し、要素名規則性特定部6によって指定される要素名(項目)を設定し、要素名規則性定義部3によって設定されたルール情報に対応するデータをレイアウト領域内から抽出し、抽出されたデータを要素関係と要素名規則性に対応づけて、データベース情報として構成要素出力部8に出力する。

【0027】以上説明した実施形態において、図5のステップA3において変換データ出力装置4の画面上には、例えば、図2に示すような商品カタログに関する情報が出力される。また、この商品カタログを構成する各要素をレイアウト領域として表示させる場合は、図9に示すように、各レイアウト領域が、矩形によって表示される。

【0028】次に、要素関係特定部5において、図7ステップB23の構成要素関係特定処理により、各要素間の関連づけが行われると、例えば、図10符号(e)に示すように、関連づけされた各レイアウト領域が強調さ

れて表示される。ここでは、1つのグループを説明するための画像と文字の情報が関連付けされている場合について図示されている。

【0029】次に、要素名規則性特定部6において、図7ステップB24の要素名規則性特定処理により、要素名に対して項目が設定されると、例えば、図10符号(f)に示すように、項目名として設定され、表示される。また、要素名規則性定義部3において、定義されたルール情報は、例えば、図10符号(g)に示すように、抽出する対象の文字のフォント、カラー、文字飾りなどについて表示される。さらに、図8のステップB241からステップB245によって構成要素名とルール情報の対応づけが行われると、例えば、図10符号(f)に示す項目に対して図10符号(h)に示すような設定がなされる。

【0030】図11は、構成要素抽出部7によって抽出されたデータがデータベース化された一例を示す図面である。構成要素抽出部7によって抽出されたデータは、関連づけされた各要素から抽出されたデータが1つの商品データとして符号(i)に示すように記憶される。また、符号(i)に示す商品データの各要素は、要素名規則性特定部6によって定義された項目名が設定され、データベースの対応する項目の欄に記憶される。以上のようにして、商品カタログなどのレイアウト構造ファイルからデータ構造ファイルを生成し、データベースを構築することができる。なお、レイアウト領域内の要素が画像である場合、データベースには、画像のファイル名が記憶される。

【0031】次に、第2の実施形態について説明する。ここでは、要素関係特定部5が行う各要素間の関連づけについての他の実施形態について説明する。この実施形態において、要素関係特定部5は、基準となるレイアウト領域である基準レイアウト領域と、基準レイアウト領域に対して相対位置条件に該当するレイアウト領域とを関連づけを行う相対位置関連づけ機能を有する。相対位置条件は、上述した入力デバイスから、作業者によって入力される。また、この相対位置条件とは、基準レイアウト領域と関連づけする対象となるレイアウト領域との位置関係を指定するための条件であり、例えば、基準となるレイアウト領域の矩形の四隅の座標からリンクされる対象となるレイアウト領域の四隅の座標までの距離及び方向を指定する情報である。

【0032】次に、この実施形態において、要素関係特定部5が相対位置に基づいて各要素間の関連づけを行う動作について説明する。まず、作業者は、関連づけを行う基準となるレイアウト領域である基準レイアウト領域を指定し、さらに、この基準レイアウト領域に対する相対位置条件を入力デバイスを介して要素関係特定部5に設定する。要素関係特定部5は、設定された基準レイアウト領域に対して相対位置条件に該当するレイアウト領

域を検索し、該当するレイアウト領域が検出された場合に、検出されたレイアウト領域と基準レイアウト領域とを関連づける。

【0033】次に、第3の実施形態について説明する。図12は、第3の実施形態における要素関係特定部5の構成を示す概略ブロック図である。この図において、相対位置情報生成部51は、基準レイアウト領域を決定し、この基準レイアウト領域に対して他のレイアウト領域までの相対位置に関する情報である相対位置情報を基準レイアウト領域以外の各レイアウト領域に対して生成する。

【0034】検出部52は、生成された相対位置情報を該当するレイアウト領域が他のレイアウト領域を基準にした場合においても存在するか否かを検出する。リンク設定部53は、検出部52によって他のレイアウト領域を基準にした場合においても相対位置情報を該当するレイアウト領域があることが検出された場合に、基準とされたレイアウト領域と相対位置情報を該当するレイアウト領域との関連づけを行う。

【0035】次に、この実施形態における要素関係特定部5の動作について説明する。ここでは、図13に示すような9種類の照明器具が掲載された商品カタログにおいて、商品毎に要素を関連付けする場合について説明する。関連付け処理の開始が上述の入力デバイスから指示されると、要素関係特定部5の相対位置情報生成部51は、符号(j)に示すレイアウト領域を基準レイアウト領域として決定し、この符号(j)の基準レイアウト領域から他のレイアウト領域までの相対位置情報を、基準レイアウト領域以外の各レイアウト領域に対して生成する。この相対位置情報の生成は、例えば、符号(k)、符号(l)、符号(m)、符号(p)、符号(r)、…等の基準レイアウト以外のレイアウト領域すべてを対象に総当たりで行われる。そして、相対位置情報生成部51は、生成した相対位置情報を一時保持する。

【0036】相対位置情報が生成されると、検出部52は、生成された相対位置情報に該当するレイアウト領域が他のレイアウト領域を基準にした場合においても存在するか否かを検出する。この検出は、例えば、符号(j)の基準レイアウト領域に対して符号(k)、符号(l)の相対位置にあるレイアウト領域は、符号(m)に示す画像が設定されたレイアウト領域を基準レイアウト領域とした場合、符号(n)、符号(s)に示すレイアウト領域が相当するので、符号(j)の基準レイアウト領域に対して符号(k)、符号(l)の相対位置にあるレイアウト領域の相対位置情報が、引き続き保持される。

【0037】一方、符号(j)の基準レイアウト領域に対して符号(r)の相対位置にあるレイアウト領域は、符号(m)に示す画像が設定されたレイアウト領域を基準レイアウト領域とした場合、相対位置が同じ場所にレ

イアウト領域が存在しないので、符号(j)の基準レイアウト領域に対して符号(r)の相対位置にあるレイアウト領域の相対位置情報が削除される。

【0038】このようにして、他のレイアウト領域を基準レイアウト領域とした場合に、当初の基準レイアウト領域からの相対位置が同じ位置にレイアウト領域が存在しなければ相対位置情報が削除され、当初の基準レイアウト領域からの相対位置が同じ位置にレイアウト領域が存在する場合に、相対位置情報が保持される。

10 【0039】そして、すべてのレイアウト領域に対して検出部52による検出処理が完了すると、リンク設定部53は、検出部52による検出処理が完了した時点において保持されている相対位置情報に基づいて、基準とされたレイアウト領域と相対位置情報に該当するレイアウト領域との関連づけを、すべてのレイアウト領域について行う。これにより、例えば、符号(j)、符号(k)、符号(l)のレイアウト領域が1つのグループとして関連付けされるとともに、符号(m)、符号(n)、符号(s)のレイアウト領域が1つのグループとして関連付けされる。さらに、他のレイアウト領域についても、符号(j)、符号(k)、符号(l)と相対位置が同じである各レイアウト領域を1つのグループとして関連付けが行われる。

20 【0040】上述の処理により、図13に示す照明器具の画像と型式と価格などについて、9種類の商品毎に、グループとして関連付けがなされる。なお、図14に示すように、符号(t)に示すレイアウト領域のみ他のレイアウト領域における関連付けが異なる場合、符号(t)に示す部分以外については、上述のグループとしての相対位置情報に基づく関連付け処理を行い、符号(t)に示す部分については、第1の実施形態において説明した、入力デバイスからの指示による関連付けを行うようにしても良い。

30 【0041】次に、第4の実施形態について説明する。この実施形態において、要素関係特定部5は、各レイアウト領域間において既に関連付けされた情報をテンプレートリンク情報として生成し、生成されたテンプレートリンク情報を他のレイアウト領域間(あるいは他のページ)において利用し、各レイアウト領域間において同様の関連付けを行う。例えば、図13に示すようにレイアウト領域が配置されたページが複数ページにわたって存在する場合に、例えば、最初のページにおいて各商品について第3の実施形態の関連付け処理を利用して各レイアウト領域間の関連づけを行い、この関連付けされた各レイアウト領域間の相対位置に基づいて、テンプレートリンク情報として生成する。そして、次のページにおいても同じレイアウト領域が配置されていれば、生成されたテンプレートリンク情報を利用して、このページ内における各商品の画像や型式、価格などの文字を1つの商品として関連付けを行う。

【0042】次に、第5の実施形態について説明する。図15は、第5の実施形態における要素関係特定部5の構成を示す概略ブロック図である。この図において、リンク情報記憶部55は、既に関連付けされた各レイアウト領域間のリンク情報を記憶する。比較部56は、対応するデータベース情報のリンク情報を読み出して、読み出されたデータの各要素と更新されたデータベース情報の各要素とを比較する。

【0043】配色部57は、比較部56の比較結果に基づいて、読み出されたデータの要素と更新されたデータベース情報の要素とが一致する要素と、一致しない要素とを異なる表現方法によって出力する。この異なる表現方法とは、例えば、異なる色をレイアウト領域に設定して出力する。また、この異なる表現方法には、レイアウト領域を示す矩形の線の種類によって区別してもよく、また、レイアウト領域にハッチングを行うなどによって区別するようにしてもよい。

【0044】リンク設定部58は、比較部56の比較結果に基づいて、読み出されたデータの要素と更新されたデータベース情報の要素とが一致する要素とを関連付けする。

【0045】次に、この実施の形態における要素関連特定装置5の動作について説明する。リンク情報記憶部55は、データベースが構築時において、関連付けされた各レイアウト領域間のリンク情報を順次記憶する。そして、データベースが構築された後あるいは構築途中において、DTPドキュメントデータの修正が行われた後に、再度関連付けを行う指示が入力されると、要素関連特定装置5の比較部56は、再度関連付けを行う指示におうじて、修正が行われたデータベース情報のリンク情報を読み出して、読み出されたデータの各要素と更新されたデータベース情報の各要素とを比較し、比較結果を配色部57に出力する。

【0046】配色部57は、比較部56の比較結果に基づいて、読み出されたデータの要素と更新されたデータベース情報の要素とが一致する要素と、一致しない要素とを異なる色をレイアウト領域に設定してリンク設定部58に出力する。リンク設定部58は、比較部56の比較結果を配色部57を介して受け取り、読み出されたデータの要素と更新されたデータベース情報の要素とが一致する要素とを関連付けする。

【0047】例えば、図16符号(u)内のレイアウト領域のうち、符号(w)に示すレイアウト領域に対してDTP上でデータの修正が行われ、上述の配色処理が行われると、図16に示すように、符号(u)内のレイアウト領域のうち、符号(w)に示すレイアウト領域が異なる色によって画面上に出力される。これにより、作業者にとって、DTP上でデータの修正が行われたレイアウト領域のデータが視覚的に確認しやすくなる。

【0048】次に、第6の実施形態について説明する。

図17は、第6の実施形態における要素関係特定部5の構成を示す概略ブロック図である。この図において、矩形内レイアウト領域検出部510は、既に関連付けされている複数のレイアウト領域を包含する最小外接矩形内のレイアウト領域を検出する。ここでいう最小外接矩形とは、例えば、図18に示すように、符号(イ)に示す画像領域と文字領域とが既に関連付けされている場合において、既に関連付けされている符号(イ)に示す画像領域と文字領域とを包含して概説する最小の矩形(符号(ホ))である。

【0049】リンク設定部520は、矩形内レイアウト領域検出部510が検出した最小外接矩形内のレイアウト領域をそれぞれ関連付けする。

【0050】次に、この実施形態における要素関係特定部5の動作について説明する。符号(イ)に示す画像領域と文字領域とが既に関連付けられており、さらに、符号(ハ)に示す図形領域についても関連付けする場合、矩形内レイアウト領域検出部510は、作業者から入力デバイスを介して、最小外接矩形による関連付け処理の指示が入力されると、既に関連付けされている複数のレイアウト領域を包含する最小外接矩形内のレイアウト領域を検出する。ここでは、符号(ホ)に示す最小外接矩形内のレイアウト領域が検出される。

【0051】最小外接矩形内のレイアウト領域が検出されると、リンク設定部520は、検出された最小外接矩形内のレイアウト領域をそれぞれ関連付けする。これにより、符号(イ)に示す画像領域と文字領域と符号(ハ)に示す図形領域が関連付けされる。

【0052】符号(ホ)内の関連付けが終了すると、他のレイアウト領域においても最小外接矩形による関連付け処理が行われる。これにより、符号(ロ)に示す画像領域と文字領域とが既に関連付けされている場合、既に関連付けされている符号(ロ)の2つのレイアウト領域を包含する最小外接矩形内のレイアウト領域(符号(ヘ))が検出され、符号(ヘ)内に存在する図形領域(符号(ニ))についても、既に関連付けされている符号(ロ)の2つのレイアウト領域に対して関連付けされる。

【0053】以上説明した実施形態によれば、印刷業者は、依頼者によって記入された見本の訂正内容に基づいて、二重に入力作業を行う必要がなくなり、作業の負担が軽減されるとともに、データベースの更新作業の時間を短縮することができる。また、作業者のミスを無くし、商品データベースに確実な更新作業を行うことができる。

【0054】また、図1における電子文書変換部1、変換データ分析部2、要素名規則性定義部3、要素関係特定部5、要素名規則性特定部6、構成要素抽出部7の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録された

13

プログラムをコンピュータシステムに読み込ませ、実行することによりデータベース構築処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものとする。

【0055】また、「コンピュータシステム」は、WW Wシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含むものとする。また上記プログラムは、前述した機能の一部を実現するためのものであっても良く、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであっても良い。

【0056】以上、この発明の実施形態を図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれる。

【0057】

【発明の効果】以上説明したように、この発明によれば、電子文書から抽出する要素を指定するためのルール情報を設定し、電子文書から抽出される要素にデータベースの項目となる項目情報を設定し、要素名規則性定義手段によって指定されるルール情報に基づいて、電子文書から要素を抽出し、抽出した要素を項目情報を対応付けてデータベース情報として出力するようにしたので、レイアウトの構造を示すファイルからデータ構造のファイルを生成してデータベースを簡単に構築することができ、作業者の負担を軽減させることができた。

【0058】また、この発明によれば、各要素間を1つのグループとして関連付けするリンク情報を生成し、リンク情報に基づいて電子文書から抽出される要素を関連付けるとともに、項目情報を対応付けてデータベース情報として出力するようにしたので、各要素をグループ毎に分類してデータベースを構築することができる効果が得られる。

【0059】また、この発明によれば、電子文書は、商品のカタログに関する情報を含むようにしたので、カタログを印刷するための要素がレイアウトされた印刷データからデータベース情報を生成することができ、これにより、カタログの情報を簡単にデータベース化することができる。

14

【図面の簡単な説明】

【図1】 この発明の一実施形態によるデータベース構築装置の構成を示す概略ブロック図である。

【図2】 変換データ表示部4に表示される画面の一例を示す図面である。

【図3】 ルール情報の一例を示す図面である。

【図4】 各要素間の関連づけについて説明するための図面である。

【図5】 図1に示すデータベース構築装置の動作について説明するためのフローチャートである。

【図6】 図1に示すデータベース構築装置の動作について説明するためのフローチャートである。

【図7】 図1に示すデータベース構築装置の動作について説明するためのフローチャートである。

【図8】 図1に示すデータベース構築装置の動作について説明するためのフローチャートである。

【図9】 商品カタログを構成する各要素をレイアウト領域として表示された場合の一例を示す図面である

【図10】 関連付けとルール情報の設定について説明するための図面である。

【図11】 構成要素抽出部7によって抽出されたデータがデータベース化された一例を示す図面である。

【図12】 第3の実施形態における要素関係特定部5の構成を示す概略ブロック図である。

【図13】 他の実施形態における関連付けについて説明するための図面である。

【図14】 他の実施形態における関連付けについて説明するための図面である。

【図15】 第5の実施形態における要素関係特定部5の構成を示す概略ブロック図である。

【図16】 第5の実施形態について説明するための図面である。

【図17】 第6の実施形態における要素関係特定部5の構成を示す概略ブロック図である。

【図18】 最小外接矩形による関連付け処理を説明するための図面である。

【符号の説明】

1 電子文書変換部	2 変換データ
分析部	

3 要素名規則性定義部	4 変換データ
表示部	

5 要素関係特定部	6 要素名規則
性特定部	

7 構成要素抽出部	8 構成要素出
力部	力部

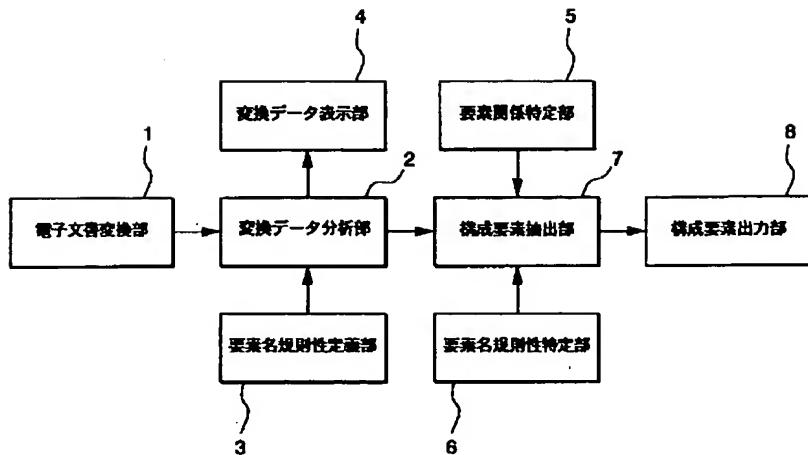
51 相対位置情報生成部	52 検出部
53、58、520 リンク設定部	

55 リンク情	
報記憶部	

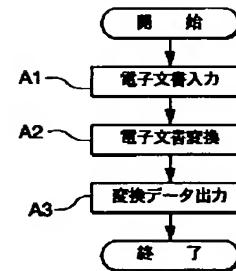
56 比較部	57 配色部

510 矩形内レイアウト領域検出部	
-------------------	--

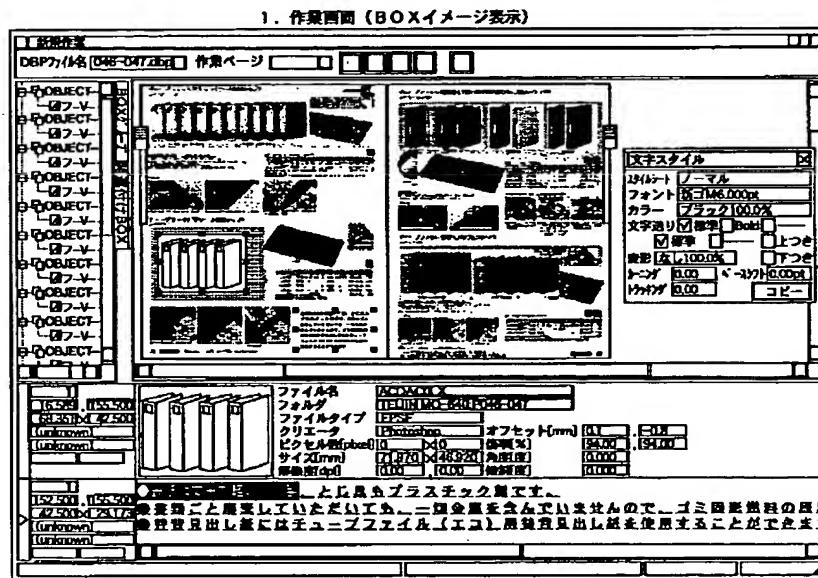
【図1】



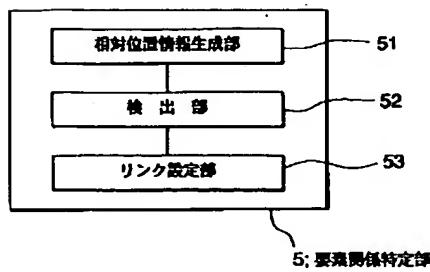
【図5】



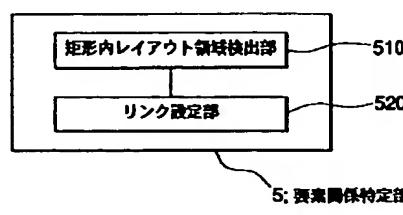
【図2】



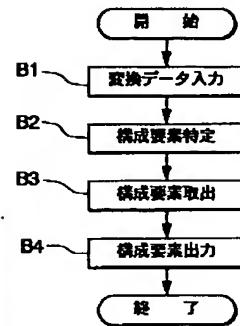
【図12】



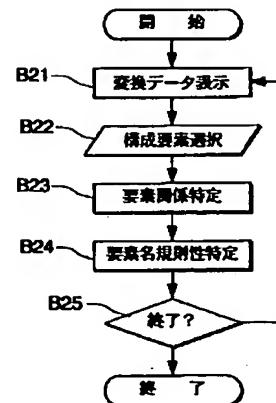
【図17】



【図6】



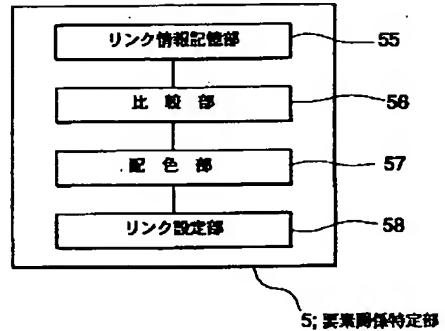
【図7】



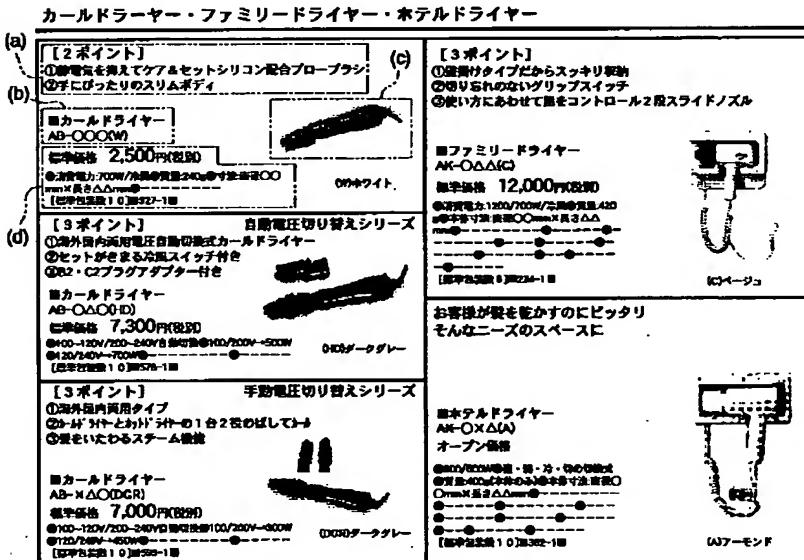
【図3】

文字の条件	文字体裁:書体名、サイズ、色、変形、字通り等 文字位置:行頭、行末等 前後関係:前後の文字や他の要素との関係 包含文字:内部に含まれる文字列
画像の条件	ファイル情報:ファイル名、ファイル種別等 画像属性:大きさ、解像度等 レイアウト属性:回転、倍率等
共通の条件	レイアウト領域:レイアウト領域の大きさ、枠線等

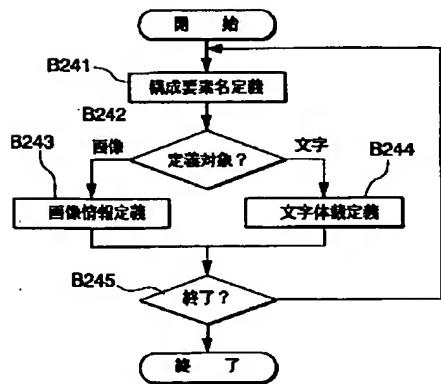
【図15】



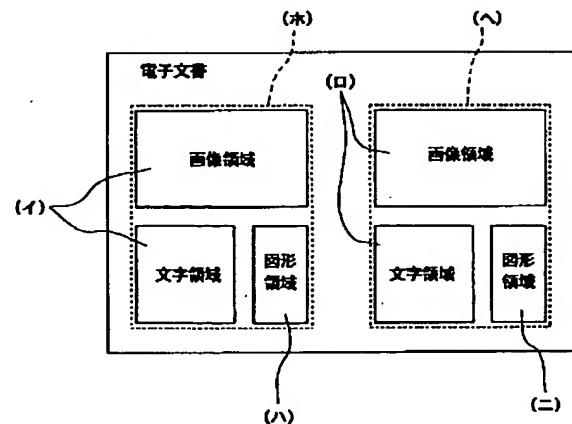
【図4】



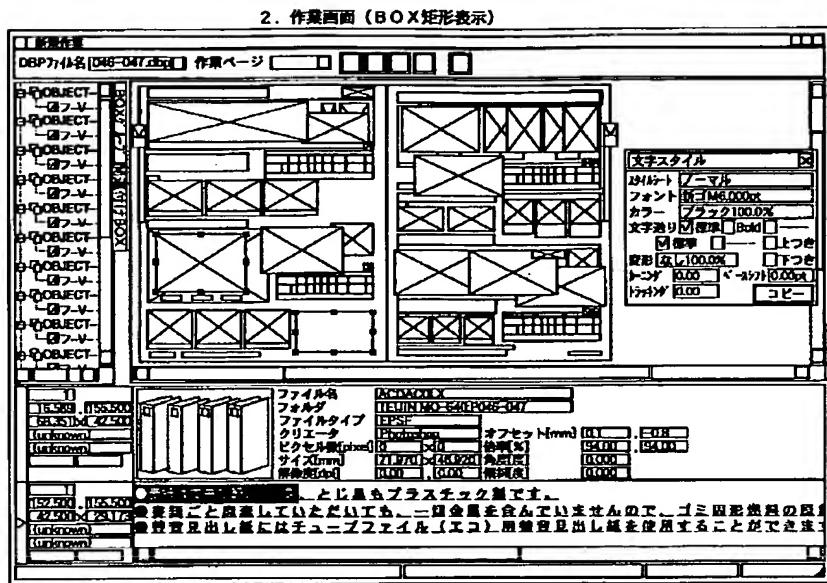
【図8】



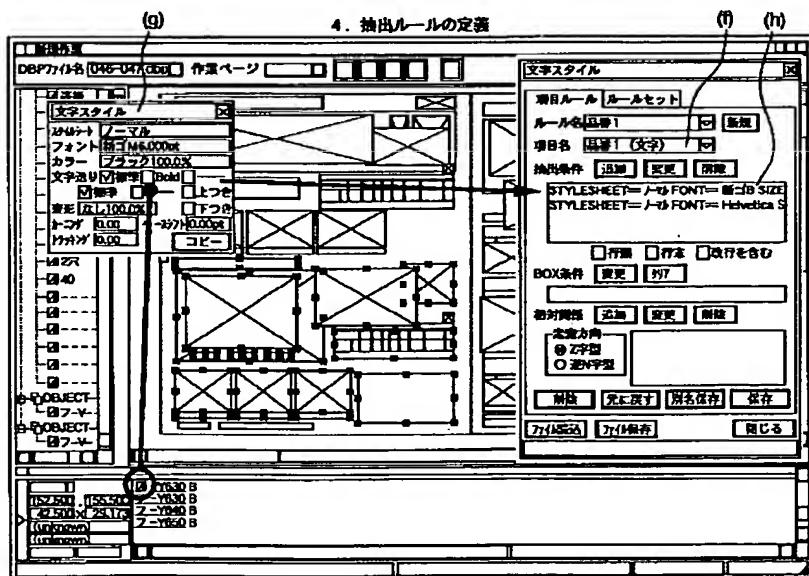
【図18】



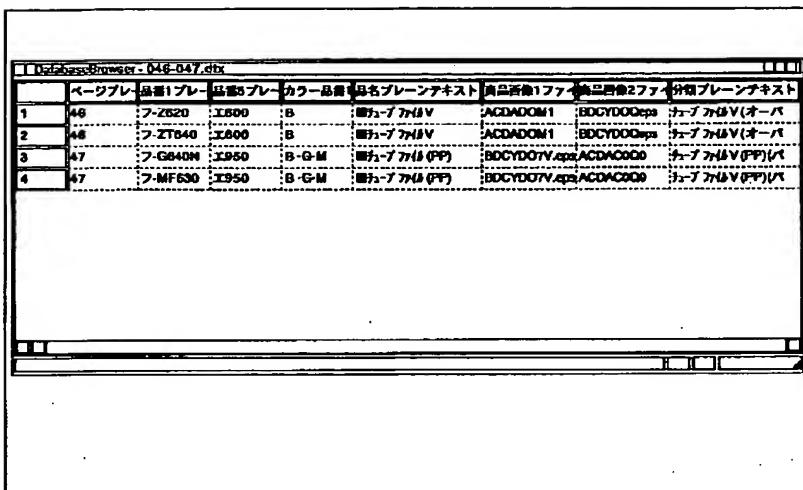
【図9】



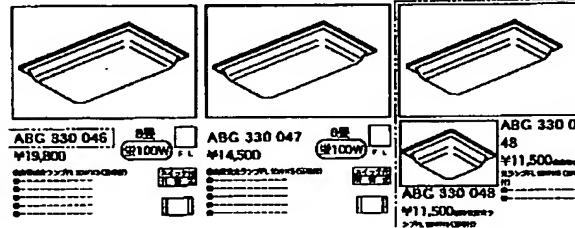
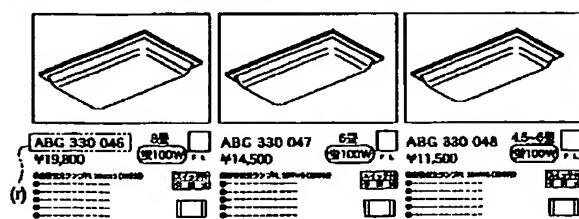
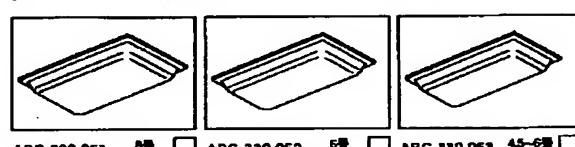
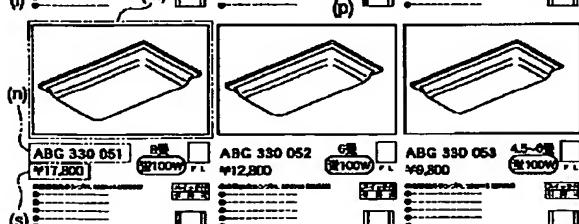
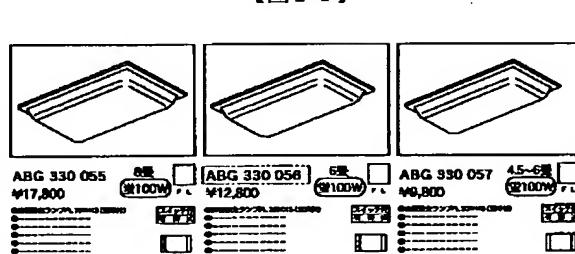
【図10】



【図11】



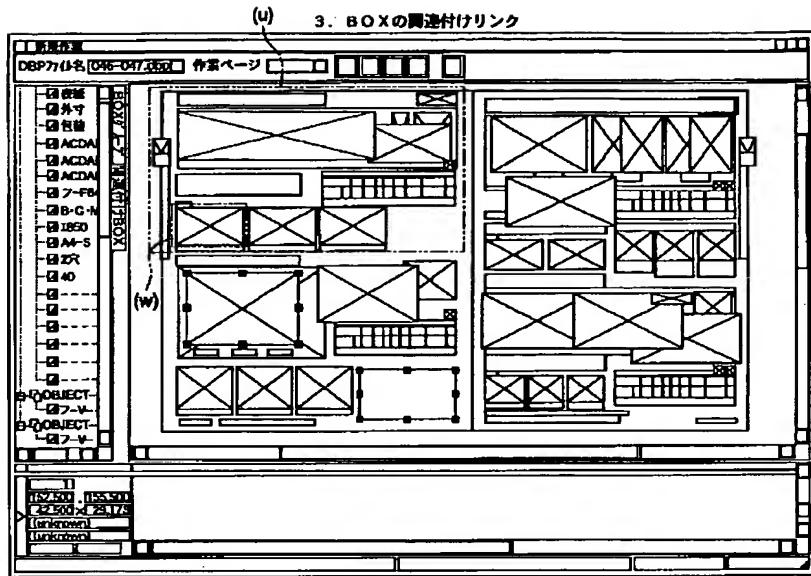
【図13】



【図14】

(l)

【図16】



フロントページの続き

(72)発明者 永山 惟

Fターム(参考) 5B082 GA07

東京都台東区台東1丁目5番1号 凸版印刷株式会社内